From Acceleration to Rhythmicity: Smartphone-Assessed Movement Predicts Properties of Music

Melanie Irrgang[a], Jochen Steffens[b], and Hauke Egermann[c]

[a]Audio Communication Group, Technische Universität Berlin, Berlin, Germany;
[b]Institute of Sound and Vibration Engineering, Düsseldorf University of Applied Sciences, Düsseldorf, Germany;
[c]York Music Psychology Group, Department of Music, University of York, York, UK

ARTICLE HISTORY
Compiled January 8, 2020

ABSTRACT
Music moves us. Yet, querying music is still a disembodied process in most music recommender scenarios. New mediation technologies like querying music by movement would take account of the empirically well founded knowledge of embodied music cognition. Thus, the goal of the presented study was to explore how movement captured by smartphone accelerometer data can be related to musical properties. Participants (N = 23, mean age = 34.6 yrs, SD = 13.7 yrs, 13 females, 10 males) moved a smartphone to 15 musical stimuli of 20s length presented in random order. Motion features related to tempo, smoothness, size, and regularity were extracted from accelerometer data to predict the musical qualities "rhythmicity", "pitch level + range" and "complexity" assessed by three music experts. Motion features selected by a stepwise AIC model predicted the musical properties to the following degrees "rhythmicity" (R^2 = .45), "pitch level and range" (R^2 = .06) and "complexity" (R^2 = .15). We conclude that (rhythmic) music properties can be predicted from the movement it evoked, and that an embodied approach to Music Information Retrieval is feasible.

KEYWORDS
Movement, Computing, Music Information Retrieval, Accelerometer, Embodied Music Cognition

1. Introduction

Music moves us. Music-induced movement ranges from spontaneous movement of head or feet to the beat, to complex dancing choreographies. Thus, music and movement share a very close relationship. There even are neuroscientific observations that listeners internally mimick body movements when perceiving music (Grahn & Rowe, 2009; Zatorre, Chen, & Penhune, 2007). That is why Maes, Leman, Palmer, and Wanderley (2014) claim that the listener's musical mind could be represented by body movement, and that the body and its movements are active contributors in musical meaning formation.

Movement is evoked by rhythmic (e.g., meter and tempo) as well as tonal or expressive qualities of music (e.g., melody, timbre, sound intensity) (Godøy, Song, Nymoen,

CONTACT M. Irrgang. Email: melanie.irrgang@tu-berlin.de

Haugen, & Jensenius, 2016; Honing, 2012; Leman & Maes, 2015; Van Dyck, Burger, & Orlandatou, 2017). Rhythmic complexity, in particular, drives listeners' desire to move and the experience of pleasure (Sioros, Miron, Davies, Gouyon, & Madison, 2014; Witek, Clarke, Wallentin, Kringelbach, & Vuust, 2014). Witek et al. (2014) confirmed an inverted U-shaped relation between complexity and urge to move. Thus, modest use of syncopation makes us move most while very simple and very complex syncopation minimizes our urge to move. Furthermore, humans prefer tempi around 120 beats per minute (bpm) to synchronize their movements to the music (Moelants, 2002). It is assumed that the musical tempo does not only affect the movement's speed but its form as well, for example a decrease in walking step size with increasing tempo (Styns, Van Noorden, Moelants, & Leman, 2007). As the speed of human movement is limited for high tempi, movement needs to be smaller or fall back to a slower metrical level (half-time). This can be observed in samba or flamenco dance.

Toiviainen, Luck, and Thompson (2010) observed that different body parts embody different levels of metric hierarchy. A follow-up study of Burger, Thompson, Luck, Saarikallio, and Toiviainen (2013) confirmed that faster metric levels induce movement on body parts featuring more degrees of freedom (arms), and slower metric levels and pulse clarity are modeled by the center of the body (torso). Movement of extremities (hand, head) further was found to be adjusted to the dynamics of the music (spectral flux of sound or loudness) (Burger et al., 2013; Camurri, Mazzarino, Ricchetti, Timmers, & Volpe, 2004; Küssner, Tidhar, Prior, & Leech-Wilkinson, 2014; Van Dyck, Moelants, Demey, Deweppe, & Leman, 2013; Wöllner & Hohagen, 2017). Starting from the above observations, Burger, London, Thompson, and Toiviainen (2018) examined the effect of spectral flux on music-induced movement under varying conditions of musical tempo. In their study, tempi were selected close to the preferred tempo of 120 bpm and ranged from 105 to 130 bpm (Burger et al., 2018). For music samples with slower tempo but strong low-frequency spectral flux in particular, they observed vertical movement of the bodily center (hips and feet) synchronizing to the beats. Extremities (hand and head) were rather synchronized to weak flux on the higher metrical bar level. Thus, these studies suggest that the center of the body is responsible to keep track of the tempo while hands or head are free to perform gestalt-like movements. Finally, dancers also respond to accents in rhythmic patterns featuring salient changes in spectral flux (Naveda, Gouyon, Guedes, & Leman, 2011). Considering the direction of movement, Burger, Thompson, Luck, Saarikallio, and Toiviainen (2014) tested to which degree music with dominant beat structure would facilitate periodic movement (Burger et al., 2014), and to which degree participants would share a common relationship between swaying movement and musical meter. Results for both hypotheses confirmed the dynamic attending theory (Drake, Jones, & Baruch, 2000; Jones, 1976) stating that vertical movement was stronger linked to beat whereas horizontal movement was stronger connected to meter.

On top of rhythm or energy-related aspects, experiments also observed interaction between pitch contour (melody) and motion in two prevalent ways. First, rising-falling pitch was shown to be embodied by movement on the vertical axis (Küssner et al., 2014; Sievers, Polansky, Casey, & Wheatley, 2013; Truslit, 1938). This observation is more consistent among musically trained persons (Küssner et al., 2014). Second, there is a strong link between pitch and size. As Eitan, Schupak, Gotler, and Marks (2014) have shown, rising pitch is associated with increasing object size whereas falling pitch is linked to shrinking object size. In experiments with free movement, it becomes even more complex: Kelkar and Jensenius (2018) instructed participants in a sound-tracing experiment to move as if they were producing the sound. Their results showed that participants apply very different movement strategies that are related to size, the vertical axis, as well as percussive elements. They furthermore observed that the direction of movement also depends on the size of the person: Smaller participants made more use of the horizontal dimension whereas taller persons remained in the vertical direction. Such effects can be the product of ergonomic aspects, social norms, or personality (Luck, Saarikallio, Burger, Thompson, & Toiviainen, 2010).

Despite these findings and ongoing research in the field of embodied music cognition, technologies in the field of Music Information Retrieval (MIR) and Music Recommender Systems (MRS) are still mainly disembodied. Leman (2007) suggests to use corporeal articulations as a bridge between linguistic self-report measures and measurements of physical energy like *pitch*, *loudness* or *tempo*. Leman conceptualizes human movement as a transformation of physical energy into cultural abstraction, and vice versa, i.e. a sonification of human movement would lead to physical energy again (Leman, 2007). He assumes that corporeal descriptions of music are more similar among humans than verbal ones, and that they are hence more appropriate and direct: Corporeal descriptions of music would close the semantic gap between physical measurements of music (audio content analysis) and subjective, verbal descriptions of music perception.

Currently, music recommendations are mostly based on previous listening habits, and the co-occurrences of songs across listeners' playlists. However, musical choices depend on situational factors, such as mood, activity-at-hand, and the presence of other people (Greb, Steffens, & Schlotz, 2018). We previously showed that the perceived emotional qualities of music (in terms of the Geneva Emotion Music Scale as well as Valence and Arousal) could be predicted based on the free and spontaneous movement it evoked during listening (Irrgang & Egermann, 2016). Since this study and others Giordano, Egermann, and Bresin (2014) showed that movement can reflect emotion, querying music by movement would open up a new, intuitive and context-sensitive way to interact with music databases without the hassle of scrolling, clicking and filling forms. In future applications, users would perform movements with a suitable input device and subsequently receive a recommendation, such that this recommendation could as well be the sonification of the very movement. In a mobile recommender scenario, suitable input devices to assess movement are smartphones, smart watches, and their inherent motion sensors. In a stationary scenario, video- (and depth sensor) based approaches like the kinect sensor are another option. The most available sensors are currently those in Android smartphone devices (market share of 85%, 2.7 billion Android devices worldwide)[1].

2. Aims

Given the close relationship between music and movement on the one side, and a lack of suitable mediation technologies on the other, the goal of the presented exploratory study was to examine whether and how gestures captured by smartphone-assessed motion data can predict musical qualities presented to listeners. In particular, we wanted to determine how smartphone-assessed accelerometer data needs to be processed and described in terms of suitable motion features to predict salient musical properties. As a step towards this goal, we asked participants to freely move a smartphone during music listening. Future applications could then recommend music featuring the pre-

[1] https://de.statista.com/themen/581/smartphones/

dicted qualities of music from movement. Thus, the findings will contribute to develop corporeal querying of music databases in MIR as envisioned by Leman (2007).

3. Method

3.1. Participants

Twenty-three persons with a mean age of 34.6 years (SD = 13.7 yrs, 13 females, 10 males) participated in this study. Six participants were professional musicians, nine were amateur musicians and eight were musical novices. Half of the participants were trained in dancing for one year or longer. Five of them were trained in dancing for three or more years (semi-professional dancers).

3.2. Stimulus selection

In order to support the musical stimulus selection, we identified several relevant musical attributes. In a study on psychophysiological responses to music, Gomez and Danuser (2007) tested the effect of ten characteristics that cover rhythmic, expressive, and tonal qualities of music. To this well-founded list, we added another four attributes (backbeat, downbeat, syncopation, beat position) related to rhythm in particular, because we expected rhythmic attributes relevant factors influencing music-evoked movement:

- rhythm (1 = vague, 10 = outstanding)
- tempo (1 = slow, 10 = fast)
- accentuation (1 = light, 10 = marcato)
- articulation (1 = staccato, 10 = legato)
- melodic direction (1 = descending, 10 = ascending)
- pitch level (1 = low, 10 = high)
- pitch range (1 = narrow, 10 = wide)
- mode (1 = minor, 10 = major)
- complexity (1 = simple, 10 = complex)
- consonance (1 = dissonant, 10 = consonant)
- backbeat (1 = vague, 10 = outstanding)
- downbeat (1 = vague, 10 = outstanding)
- syncopation (1 = accent on beat, 10 = accent on off-beat)
- beat position of bass/snare (1 = laid back, 10 = up front)

According to these musical attributes, we selected a total of 31 musical excerpts (20 seconds long), such that firstly their musical characteristics would not change (much) over time, and that secondly we created sufficient variance for each musical attribute. In a pre-study, this set of 31 musical excerpts was then rated according to the list of musical attributes by three experts from different fields of musical expertise: a music researcher, a professional musician (bass guitar), and a professional DJ. This larger set of 31 stimuli had to be reduced to a set that was large enough to represent different degrees of the above characteristics, but small enough to avoid that this physically intense study was not too long and exhausting. We estimated that participants will listen to each music excerpt at least twice before recording it in order to get to know the song, and to develop a movement strategy. For the short interview after each stimuli, we calculated another 5 minutes. Thus, 15 stimuli were identified as the optimal number

Table 1. List of music stimuli used in the study.

Artist	Title	Time Excerpt	BPM
Alicia Keys feat. Nicki Minaj	Girl on Fire (Inferno Version)	00:00-00:20	93
Marcus Miller	Detroit	00:00-00:21	92
Sia	Chandelier	00:30-00:55	117
Alicia Keys	You don't know my name	00:33-00:57	84
Fever Ray	Dry and Dusty	01:20-01:42	91
David Bowie	Aladdin Sane	02:00-02:30	121
Stevie Wonder	Another Star	02:21-03:00	122
Andy Allo	People Pleaser	00:18-00:40	85
Michael Jackson	Bad	00:00-00:20	114
Röyksopp feat. Robyn	Monument (The Inevitable End Version)	00:40-01:02	93
Igor Stravinsky	Le Sacre du Printemps/Part1	00:20-00:40	82-88
Chris Garneau	The Leaving Song	00:00-00:22	54
Jherek Bischoff & Amanda Palmer feat. Neil Gaiman	Space Oddity	01:36-02:00	130
David Bowie feat. Tina Turner	Tonight	00:18-00:48	98
Florence + the machine	Cosmic Love	00:20-00:43	134

of stimuli to remain in the scope of two hours for the experiment. In order to select 15 stimuli representing all musical characteristics in the larger set of 31 pieces, we computed k-means clustering for 15 clusters from the expert ratings and selected one stimulus from each cluster. Table 1 depicts the final list of samples which were presented in random order in the main study.

3.3. Procedure

In the beginning, participants were asked for their written consent to participate. Their motions were tracked by an Optitrack[2] motion capture system in order to have a reference measure for the accelerometer data (not shown here). Participants wore a motion capture suit equipped with 37 markers and holding the smartphone equipped with another three marker points as rigid body. Markers were tracked by eight cameras. A video camera recorded the participants' movements in order to disambiguate occluded marker points and to record short interviews about the music after each excerpt. Music was presented to participants via loudspeakers. An Android App was developed to capture motion and present music stimuli in a random order. The App was controlled by the investigator via remote access with AirDroid[3]. The whole study was conducted using a Samsung Galaxy S6. A sling around the wrist served as a safety measure for the phone not to be slipped. Except for one person, participants held the phone in the right hand because they were right-handed. Note that the phone's front surface was facing to the torso's side. Figure 1 shows the smartphone's position relative to the body.

[2]www.optitrack.com
[3]www.airdroid.com

Figure 1. Photo of smartphone's position in participant's hand indicating accelerometer dimensions X,Y,Z

In the beginning, participants chose one of their own songs to warm up and to get familiar with the study's procedure. This song could be any song that they brought and that they liked for dancing. It was neither part of the 15 stimuli nor the evaluation, and should simply ensure that they had a comfortable start. They were instructed as follows: "Move the smartphone to the music. You can move the rest of the body intuitively along with it, but keep in mind that the characteristic motion must be captured by the phone. Please stay in the delineated area of 2x2 metres". Subsequently, for every other music excerpt, the procedure was as follows:

(1) multiple test listening to the music excerpt in order to develop a movement strategy
(2) accelerometer recording of the movement during stimulus presentation
(3) short interview about the music excerpt
    (a) How much did you like moving/dancing to the song?
    (b) How much did you like the song?

Participants were allowed to listen to each stimulus as often as it took them to decide how they wanted to move to it (multiple test listening). During test listening, we did not record any data. Once they were ready, we started the main recording of movement data for the given stimulus. After the recording, we had a short interview about how suitable they considered the musical excerpt for moving/dancing to it (suitability to move), and about how much they liked the presented music in general (liking of song). They rated each song as either "not suitable"/"dislike", "moderately suitable/moderate liking", or "very suitable/strong liking". Subsequent to the study's main part, participants were asked to fill out a socio-biographic questionnaire on their experience in music and movement/dance etc.

4. Results

The following sections summarize the results for the extracted music and motion features as well as for the statistical models fitted for predicting musical features based on motion features.

original musical property	**rhythmicity**	**pitch level + range**	**complexity**
rhythm	**.912**	-.096	.226
backbeat	**.718**	.133	-.605
tempo	**.457**	.168	**.631**
accentuation	**.918**	.110	.175
articulation	**-.888**	.219	-.127
pitch level	-.138	**.914**	.088
pitch range	.044	**.895**	.252
complexity	.169	.380	**.829**

Table 2. Loadings between components and musical attributes. Components were computed via PCA and rotated (Varimax with Kaiser normalization)

4.1. Music Features

For some music features, inter-rater reliability was not sufficient, namely for syncopation, downbeat, beat position, consonance, mode, and melodic direction. Therefore, in the first step, we only kept those properties for which the inter-rater reliability was acceptable (inter-rater correlation $r > .50$). Since ratings of the remaining music properties were highly correlated as visualized in Figure 2 (e.g., *rhythm* with *accentuation* or *pitch level* with *pitch range*), we calculated a Principal Component Analysis (PCA) in a second step. Both Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO = .589) and Bartlett's Test of Sphericity indicated a (mediocre) adequacy of the sample for computing a PCA (chi-squared = 178.4, df = 28, $p < .001$). A scree plot indicated a 3-component solution that was rotated using orthogonal Varimax rotation with Kaiser normalization. The resulting components were labeled *rhythmicity*, *pitch level + range* and *complexity* (see Table 2). We then extracted component scores and computed the mean score per excerpt (across the three expert ratings). Figure 3 shows principle component means and standard deviations (rhythmicity, pitch, complexity) for all 15 music excerpts (left plot). The right plot of Figure 3 shows participant rating means and standard deviations for all 15 music excerpts on how suitable they considered the song for moving/dancing (solid line), and how much they liked the song (dashed line). Comparing the two plots, the music property most related to movement suitability is rhythmicity, e.g. highest ratings of rhythmicity and suitability to move for *Girl on Fire*, *Detroit*, *People Pleaser* or *Bad*, while lowest ratings for rhythmicity and suitability to move are present for *Leaving Song* or *Cosmic Love*. Accordingly, the correlation between *suitability to move* and *rhythmicity* is significantly positive (r = .37, $p < .001$); between *suitability to move* and *pitch level + range*, it is significantly negative (r = -.24, $p < .001$); but between *suitability to move* and *complexity* it is not significant (r = .01, p = .87). The right plot of Figure 3 further suggests that *liking* is independent from *suitability to move*. In fact, *liking* and *suitability to move* are only slightly positively correlated (r = .58, $p < .0001$).

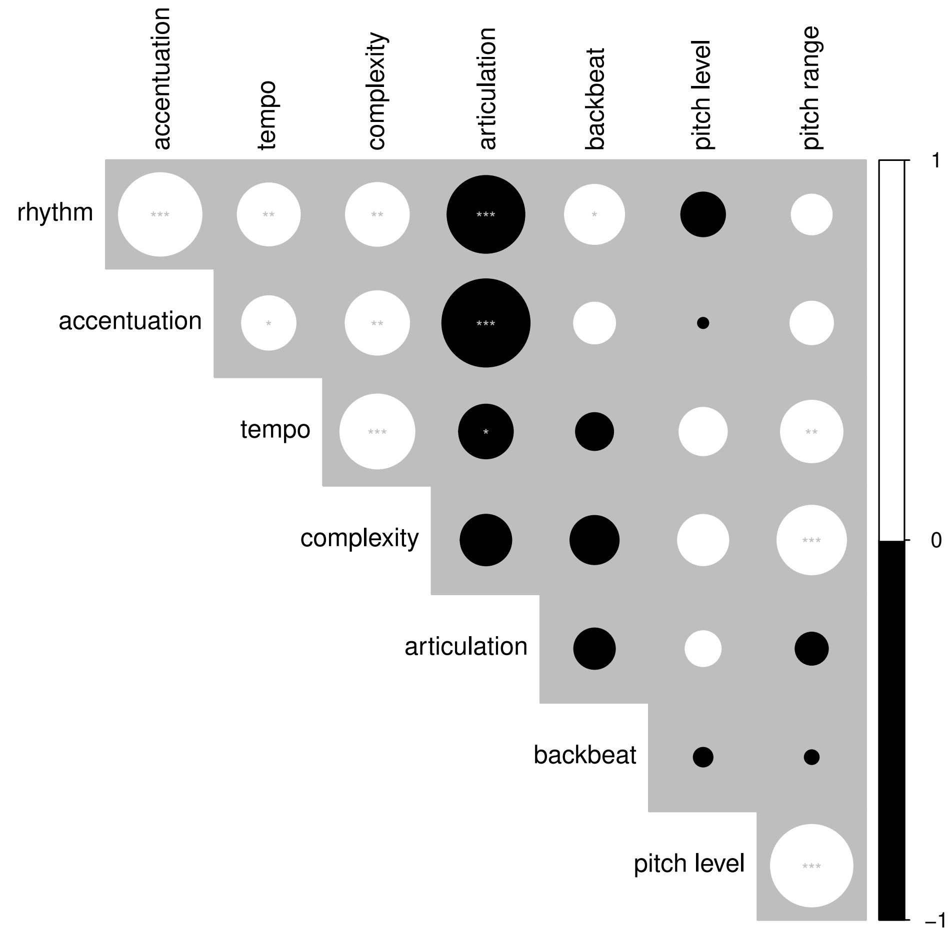

Figure 2. Correlation between music features with acceptable inter-rater reliability. The highest correlations can be observed between accentuation and articulation, between rhythm and accentuation, as well as between pitch level and pitch range. ***$p < .001$, ** $p < .01$, *$p < .05$

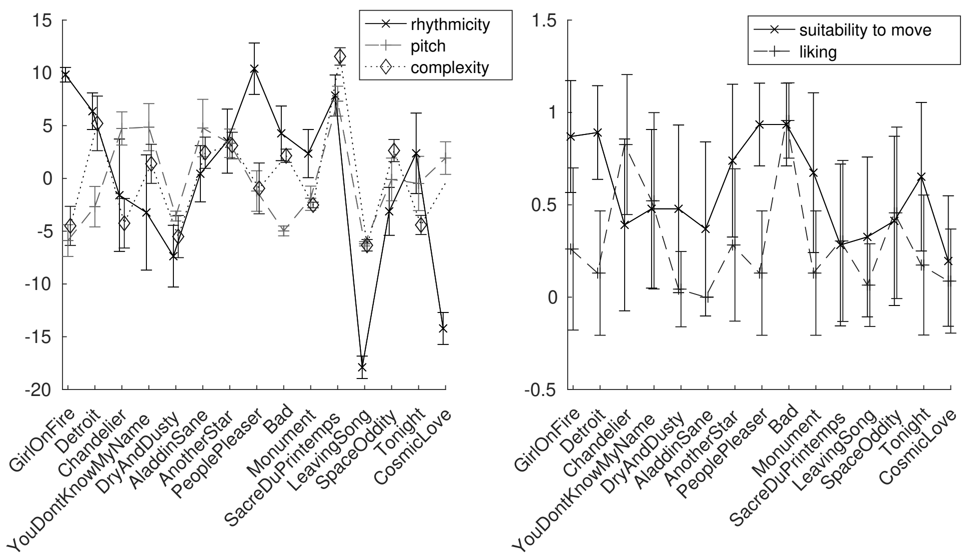

Figure 3. Mean (SD) expert rating components representing music properties (left) and participant mean (SD) rating on suitability to move and liking (right), separated by music excerpt. Comparing the two plots, the music property most related to movement suitability is rhythmicity, e.g. highest ratings of rhythmicity and suitability to move for *Girl on Fire*, *Detroit*, *People Pleaser* or *Bad*, while lowest ratings for rhythmicity and suitability to move are present for *Leaving Song* or *Cosmic Love.* The right plot also shows that liking appears to be independent from suitability to move.

4.2. Motion Features

During the experiment, some participants skipped certain songs because they did not feel comfortable with them. Those samples (N = 23) were excluded from the original set of N = 345 samples leading to a final size of N = 322 observations within 23 persons. In a pre-processing step, we standardized sample rates of smartphone-generated accelerometer data[4] to the lowest one that was captured of approximately seven samples per second. We also cut off the first five seconds of each trial because participants used this time to perform a hand signal to indicate the start for the post-processing of the data[5]. Subsequently, the pre-processed data was transformed into the eigenspace (described by eigenvectors and eigenvalues of the data) of each participant: For all samples of one participant, we computed a Principal Component Analysis (PCA) and transformed the participant's data to decorrelate it. This step increases the comparability of movements across participants. Figure 4 visualizes the steps described above for one participant and their movement to one music excerpt (AladdinSane). The first principal component (PC1) corresponds to the direction of maximum movement variance in space. The second principal component (PC2) corresponds to the direction of second most movement variance in space and so on. These directions do not correspond to the global directions of x, y, and z, because they are individual for each participant depending on how this person held the device, and based on their individual movement pattern.

From the accelerometer data in eigenspace, we extracted spectral and temporal motion features related to tempo, size, regularity and smoothness as listed in Table 3.

[4] corrected linear acceleration as described for Sensor.TYPE_LINEAR_ACCELERATION in: https://developer.android.com/guide/topics/sensors/sensors_motion

[5] They stretched out their arms to the side (shoulder's height), and then clapped them together above their head.

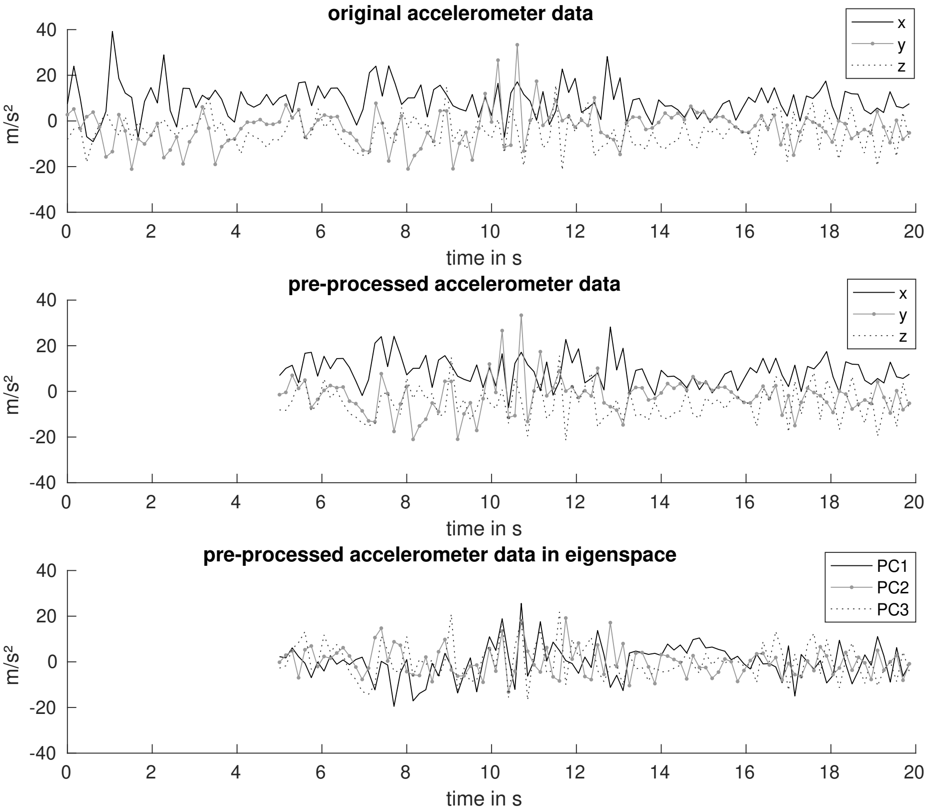

Figure 4. Recording of motion data for one musical excerpt *Aladdin Sane* and one participant. The top plot shows the original accelerometer data. In the middle, the beginning 5s were cut off and sample rate was adjusted to 7 data points per second. The bottom plot shows the participant's data in eigenspace (after PCA).

Spectral features (e.g. *maximum frequency in Hz*) were computed by applying a fast fourier transform (FFT). Temporal features (e.g. *distance between midcrosses*, *rise and fall times*) were aggregated over time by computing median and standard deviation for each feature and principal component. In order to further reduce the space of motion features, we computed the median over all three components for each feature. For example, the features peak_median_PC1, peak_median_PC2, peak_median_PC3 were reduced to median_peak_median, and peak_std_PC1, peak_std_PC2, peak_std_PC3 became median_peak_std. For spectral features, we kept the most dominant frequency (max_freq_hz), its magnitude (max_freq_mag), and its magnitude relative to the median magnitude of all frequency peaks (max_freq_rel).

Figure 5 illustrates the steps described above. This feature space, FEATURE SPACE A, comprises 12 features. Figure 6 visualizes some of the temporal and spectral features for one participant and the music excerpt *AladdinSane*. Subsequently, we calculated different measures of multicollinearity (individual multicollinearity diagnostic measures (imcdiag) from R-package mctest[6] including Variance Inflation Factor (VIF), TOL(erance) and Farrar F-test) for each regressor in order to check if the influence of a predictor variable in the presence of another dependent or correlated variable might be obscured. Collinearity was detected by this test for the variables: max_freq_hz, max_freq_mag, median_dist_midcrosses, median_rise_median, median_fall_median, median_fall_std, median_peak_std. Figure 7 visualizes the measured correlation between features of Feature Space A. Thus, in a second step, we evaluated if results remained comparable for an even smaller, decorrelated feature space. We applied a PCA on Feature Space A and only kept the first three components (criterion: scree plot). Subsequently, components were rotated according to the varimax method. This feature space, Feature Space B, comprises the dimensions *irregular slowness*, *irregular size* and *irregular smoothness*. Unfortunately, it was not possible to find a solution with regularity as a separate component to learn how regularity affects the prediction of musical properties independently of the other motion features slowness, size, and smoothness. Table 4 shows the component loadings for the three components. The following section will describe the selection of features for both feature spaces and the results of the fitted models.

[6]https://cran.r-project.org/web/packages/mctest/

feature	description	category	Matlab function
spectral			(Signal Processing Toolbox)
max_freq_hz	most dominant frequency in Hz across all three PCs	tempo	fft and max
max_freq_mag	magnitude of most dominant frequency	regularity	fft and max
max_freq_rel	magnitude of most dominant frequency relative to median magnitude of all frequency peaks	regularity	fft, findpeaks, and median
temporal			
volume	volume of acceleration point cloud computed by applying delaunay triangulation to the acceleration data points in 3D-eigenspace and hence computing the convex hull of the triangulated space as in Amelynck et al. (2012)	size	delaunayTriangulation and convexHull
dist_midcrosses_median_PC1/2/3	median of distances between midcrosses for each PC	tempo	midcross and median
dist_midcrosses_std_PC1/2/3	standard deviation of distances between midcrosses for each PC	regularity	midcross and std
rise_median_PC1/2/3	median duration of attacks for each PC	smoothness	risetime and median
rise_std_PC1/2/3	standard deviation of duration of attacks for each PC	regularity	risetime and std
fall_median_PC1/2/3	median duration of releases for each PC	smoothness	falltime and median
fall_std_PC1/2/3	standard deviation of duration of releases for each PC	regularity	falltime and std
peak_median_PC1/2/3	median size of the movement for each PC	size	peak2peak and median
peak_std_PC1/2/3	standard deviation of the size of the movement for each PC	regularity	peak2peak and std

Table 3. Feature Space A - Motion Features extracted with Matlab's Signal Processing Toolbox. The Matlab code can be retrieved from this public repository: https://github.com/mirrgang/motion2music

	irregular slowness	irregular size	irregular smoothness
max_freq_hz	**-.429**	.188	-.234
max_freq_mag_rel	**.823**	-.082	-.094
median_dist_midcrosses	**.669**	.014	-.300
median_std_dist_midcrosses	**.726**	-.085	.070
median_rise_median	**.766**	.055	.331
median_fall_median	**.876**	.051	.162
max_freq_mag	.168	**.824**	.013
volume	-.096	**.871**	.001
median_peak_median	-.272	**.925**	.033
median_peak_std	-.187	**.936**	.006
median_rise_std	-.010	.001	**.734**
median_fall_std	.140	.005	**.750**

Table 4. Loadings of varimax rotated principal components. Low and irregular movement frequencies, long rise and fall times constitute the first principal component: irregular slowness. The second component, irregular size, is composed of size-related motion feature e.g. volume or peak amplitude. The third principal component features irregular smoothness.

4.3. Statistical Prediction Models

For all three music components *rhythmicity*, *pitch level + range* and *complexity*, and both feature spaces, we computed a stepwise model based an the Akaike Information Criterion (AIC) choosing the model with minimum information loss. We further restricted selected variables to achieve the significance level of ($p < .05$), and therefore increased the penalty term from $k = 2$ (default) to the chi-squared quantile of $k = qchisq(0.05, 1, lower.tail = FALSE) = 3.84$ for the stepAIC function (R-package MASS). Stepwise search was performed in both directions[7].

Table 5 shows the results of the selected fixed effects from Feature Space A for the three factors *rhythmicity* ($R^2 = .45$), *pitch level + range* ($R^2 = .06$) and *complexity* ($R^2 = .15$). For *rhythmicity*, about half of the variance in the data could be explained by the fitted model. It was predicted by motion features related to irregular (max_freq_mag_rel and std_dist_midcrosses), large (median_peak), and irregularly sharp (std_rise) movement with dense use of space (volume). *Pitch level + range* was the most difficult factor to predict from the extracted motion features. Significant features were connected to irregularly slow (max_freq_mag_rel, max_freq_hz) movements with short rise times (median_rise). About one eighth of the variance in *complexity* was explained by features associated with irregular tempo (max_freq_mag_rel), size (std_peak) and fall times (std_fall) but regular rise times (std_rise).

Table 4.3 shows the results of the selected fixed effects from Feature Space B for the three factors *rhythmicity* ($R^2 = .36$), *pitch level + range* ($R^2 = .03$) and *complexity* ($R^2 = .09$). Like for Feature Space A rhythmicity was predicted by large movement (irregular size comprising volume and median_peak). Regularity and fast tempo (negative irregular slowness comprising max_freq_mag_rel and std_dist_midcrosses) was also a significant predictor. The simplified Feature Space B explained about 10% less of the variance in the data than Feature Space A. However, it could be shown that (fast) tempo and size explain most of the variance in the data. *Pitch level + range* was predicted by the opposite of the irregular slowness component which comprised significant features of Feature Space A. As above, *complexity* was associated with *irregular size* (cf. std_peak), as well as the opposite of *irregular slowness* (cf. max_freq_mag_rel).

[7]The complete R code is provided in the following repository: https://github.com/mirrgang/motion2music

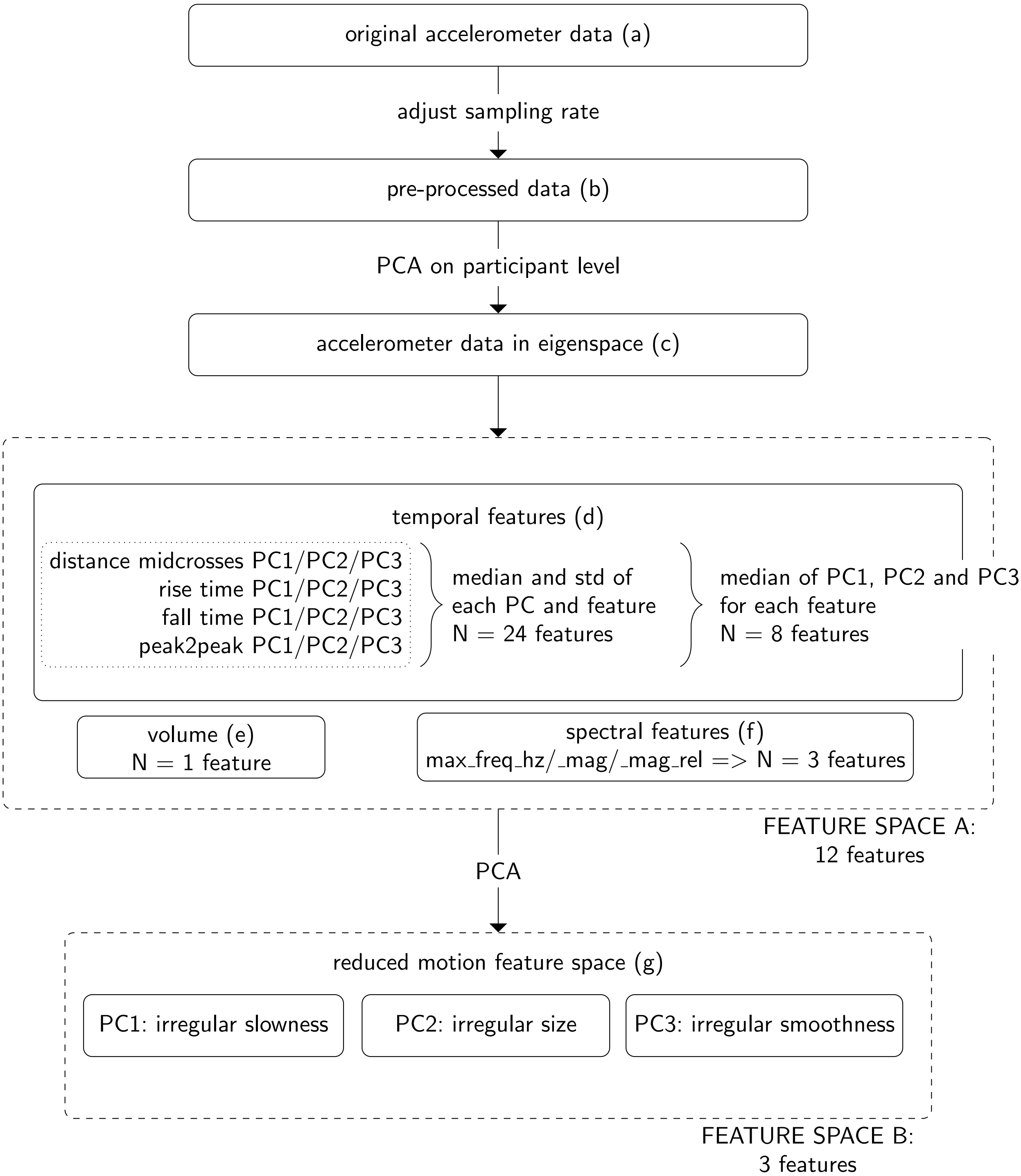

Figure 5. Motion Feature Extraction: Workflow. The original accelerometer data (a) is pre-processed (b), and a PCA is applied on participant level to account for inter-individual differences (c). Temporal features are extracted for each dimension in eigenspace (d). Median and standard deviation are computed for each feature and each component to get a time-averaged representation (N=24 temporal features). Finally, the median is computed over all components of a feature (N=8 temporal features). Volume is a measure of use of space over the whole time (e). Spectral features (max_freq_hz, max_freq_mag, max_freq_mag_rel) are assessed once for the complete time segment (f). A second PCA on *Feature Space A* yielded three components: irregular slowness, irregular size, irregular smoothness (g).

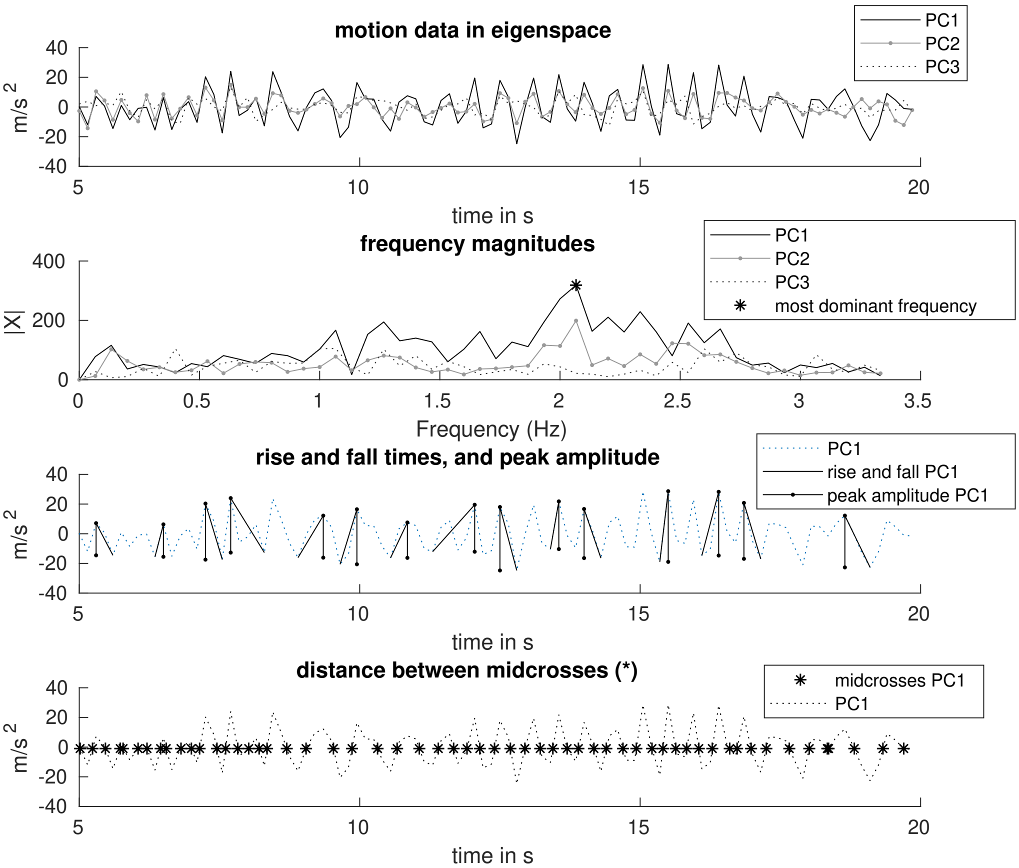

Figure 6. Sample motion features of *AladdinSane* from one participant. The most dominant frequency is in the mid spectrum around 2Hz. Acceleration is relatively regular with almost equidistant midcrossings, large peak amplitudes in the first principal component and short rise and fall times.

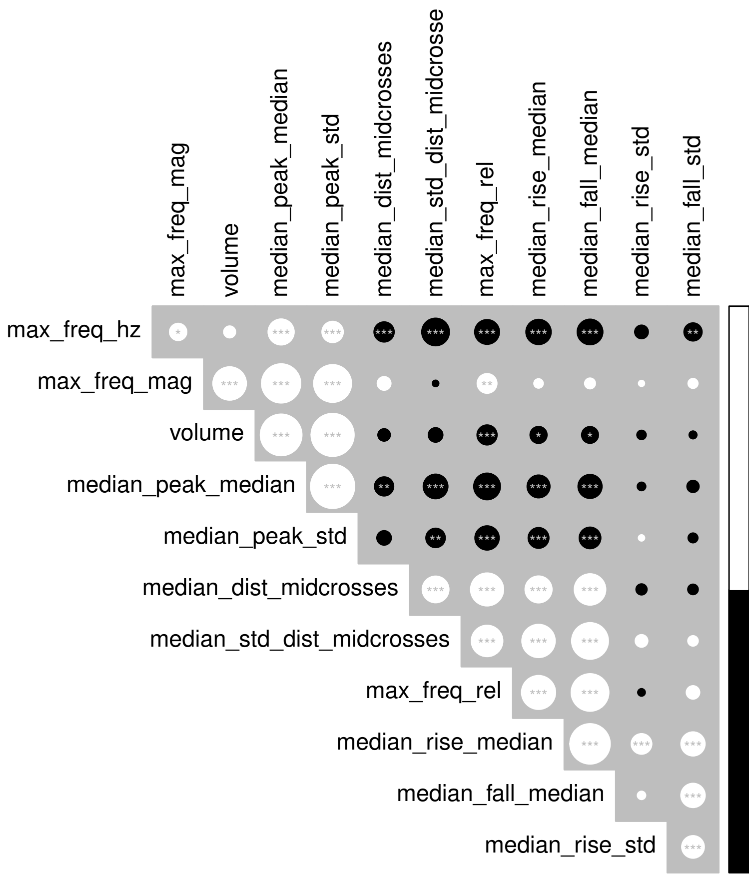

Figure 7. Correlation of Motion Features from Feature Space A. Most of the motion features are significantly correlated. *** $p < .001$, ** $p < .01$, * $p < 0.05$

Different from the model for Feature Space A, irregular smoothness (std_rise and std_fall) was not significant. The third principal component of the feature space, *irregular smoothness*, was not selected as a significant predictor in any of the three models.

Finally, we also tested the discriminative power of the extracted motion features by computing three linear mixed models that predict the motion feature components *irregular_slowness*, *irregular_size*, and *irregular_smoothness* with *sample_id* as fixed effects and *participant_id* as random intercept. The mixed models were fit using the R-package nlme[8]. Figure 8 shows the corresponding estimated marginal means and the confidence intervals for the three mixed models. The joint significance (average significance over all music excerpts) of the fixed effect *sample_id* was smaller then .01% for each model. Thus, the results indicate firstly that the different music excerpts evoke different movement features, and secondly that the variance of the music samples is suitably represented by the quantified movement.

[8]https://cran.r-project.org/web/packages/nlme

Table 5. Results from stepAIC (fixed effect estimates, t-value and p-value), and R for the fitted linear models and Feature Space A.

	name	**category**	**estimate**	**t value (p^1)**
rhythmicity ($R^2 = 0.45$)***				
magnitude of most dominant frequency relative to median magnitude of all frequency peaks	max_freq_mag_rel	regularity	-2.7	-6.9 (***)
volume of acceleration point cloud	volume	use of space	-2.0	-3.5 (.0005)
median standard deviation of distance between midcrosses	median_std_dist_midcrosses	regularity	-1.3	-3.3 (.0009)
median standard deviation of rise time over all components	median_rise_std	irregularity	-1.1	-3.4 (.0007)
median peak amplitude over all components	median_peak_median	size	4.1	6.9 (***)
pitch level and range ($R^2 = 0.06$, $\boldsymbol{p} = \mathbf{.0002}$)				
maximum frequency in hz	max_freq_hz	tempo	-0.5	-2.2 (.0256)
magnitude of most dominant frequency relative to median magnitude of all frequency peaks	max_freq_mag_rel	regularity	-0.6	-2.3 (.0228)
median rise time over all components	median_rise_median	smoothness	-0.6	-2.1 (.0330)
complexity ($R^2 = 0.15$)***				
magnitude of most dominant frequency relative to median magnitude of all frequency peaks	max_freq_mag_rel	regularity	-1.0	-4.1 (***)
median std of rise time over all components	median_rise_std	irregularity	-0.9	-3.8 (.0002)
median std of fall time over all components	median_fall_std	irregularity	0.5	2.0 (.0389)
std of peak amplitude over all components	median_peak_std	irregularity	0.9	3.8 (.0002)

1) $***p < .0001$

Table 6. Results from stepAIC (fixed effect estimates, t-value and p-value), and R for the fitted linear models and Feature Space B.

	category	**estimate**	**t value (p^1)**
rhythmicity ($R^2 = 0.36$)***			
irregular slowness	irregularity & tempo	-1.1	-9.7***
irregular size	irregularity & size	0.6	4.8***
pitch level and range ($R^2 = 0.03$, p **= .0041**			
irregular slowness	irregularity & tempo	-0.2	-2.9 (.0041)
complexity ($R^2 = 0.09$)***			
irregular slowness	irregularity & tempo	-0.2	-3.0 (.0037)
irregular size	irregularity & size	0.3	3.2 (.0014)

1) $***p < .0001$

5. Discussion

The presented study aimed at predicting properties of music based on smartphone-assessed movement. In a music recommender scenario, music featuring the predicted properties could be suggested to the user, and accordingly querying music by movement, as envisioned by Leman (2007), would provide an alternative to disembodied options of music recommendation. In order to approach this goal, we evaluated which properties of music can be predicted by motion features extracted from smartphone accelerometer data. In general, it was possible to predict musical properties by the quantified movement. However, the results varied highly between the properties *rhythmicity* (which was predicted better), *pitch level + range, and complexity* (which where predicted worse). Another open question is the necessary amount of motion features. Though, there were high correlations between variables in Feature Space A and hence redundancy, we could not explain the variance in the data to the same degree with the reduced Feature Space B (e.g. for rhythmicity $R^2_A = .45$ vs. $R^2_B = .36$). The stepAIC method that we chose for the regression models of Feature Space A selects the features with the highest predictive power. Considering rhythmicity, stepAIC still selected correlated motion features like max_freq_mag_rel and median_rise_std. Though correlated, those motion features seem to explain different variances in the data. The worse performance of Feature Space B hence suggests that a PCA might be unsuitable to reduce the amount of correlated features because the predictive potential decreased noticeably. There also is potential for motion features explaining the variance in the data that we could not explain by this approach. For future work, it might therefore be advantageous to compile a larger set of motion features, and instead of computing a PCA, stepwiseAIC or VIF selection should be applied to reduce correlated motion features.

For Feature Space B, it was also not possible to isolate a component related only to regularity. This made it difficult to interpret whether one of the predicted music properties was related to tempo or regularity. This correlation of motion variables related

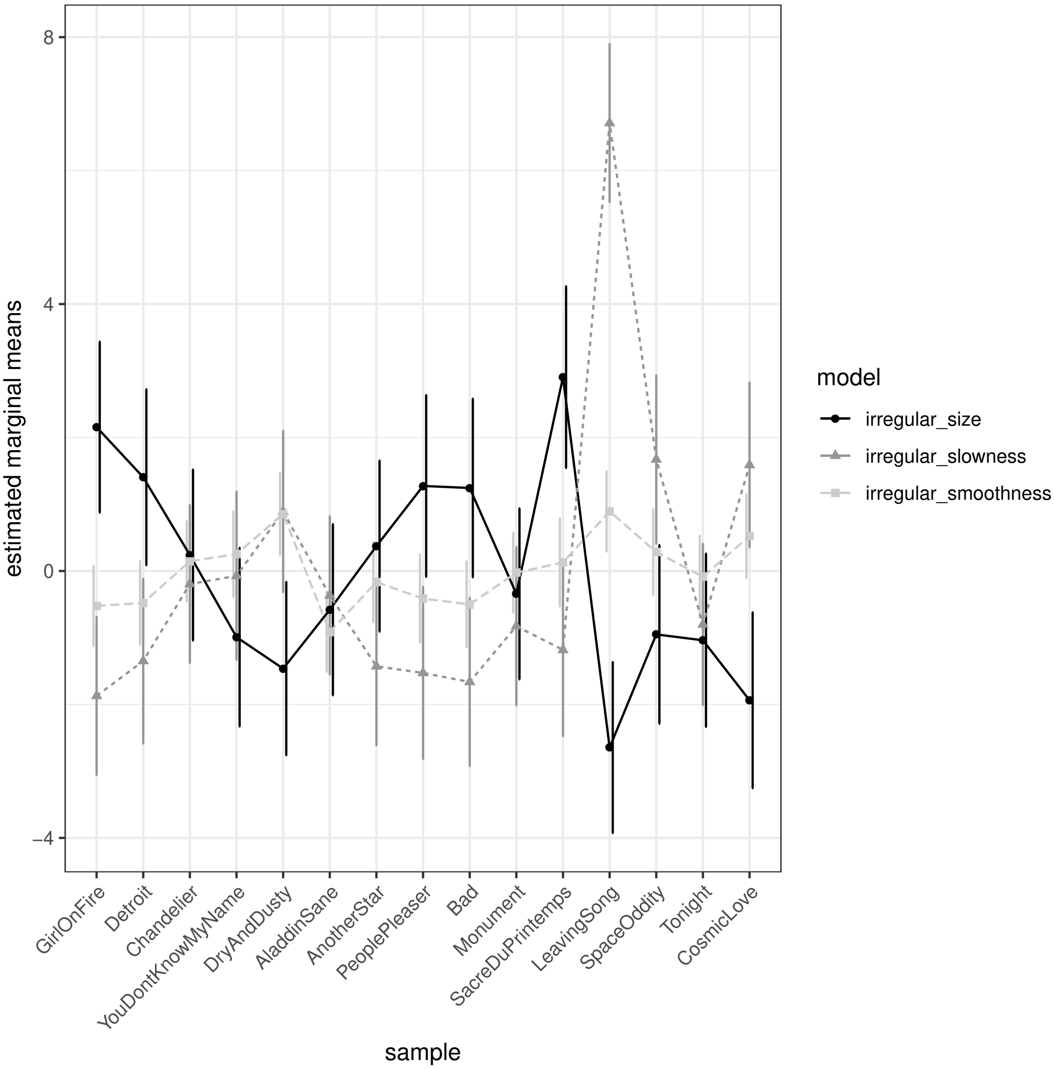

Figure 8. Estimated marginal means and confidence intervals of linear mixed models with *sample_id* as fixed effect and *irregular_slowness*, *irregular_size*, *irregular_smoothness* as dependent variables.

to regularity and tempo might also be due to the possibility that participants were only moved by music that featured enough rhythmic regularity and tempo. The majority of participants consisted of persons that were not trained in dancing, and thus just stood still to music that either was too slow (e.g. "Leaving Song") or rhythmically too complex (e.g. "Le Sacre du Printemps"). Given the potentially different movement preferences of dancers and non-dancers, it might be recommendable to have two different models for those two groups. However, we did not have enough trained dancers to estimate a separate model for them but we will consider this for future work.

In general, the negative relationship between *rhythmicity* and *irregular slowness* is in line with previous research showing that participants adjust their movements (or an object's movement) to the tempo of the music (Küssner et al., 2014; Moelants, 2002; Sievers et al., 2013). The significance of movement size for *rhythmicity* is also congruent with the results from (Witek et al., 2014) who confirmed that rhythmicity drives listeners' desire to move. Rhythm-related properties of music probably evoked more similar and repetitive movement among participants. Thus, the movement was suitably quantified to predict rhythmicity by motion features that were averaged over time.

Regarding *pitch level + range*, two metaphors are prevalent in the literature: The vertical and the size metaphor. The vertical metaphor (rising pitch = rising movement, falling pitch = falling movement) cannot be assessed by accelerometer data since it does not provide absolute position in space. The size metaphor (rising pitch = increasing object size, falling pitch = shrinking object size) could have generated a relation between the pitch property and motion features from the category size. However, for none of our two pitch models, motion features related to size like *volume* or *median peak amplitude*, or *irregular size* respectively, were estimated as significant predictors. High *pitch + range* was associated with slow tempo and sharpness in this study. Given the low predictive power for pitch, another possible explanation might be that participants preferred rhythmic over melodic qualities of music to move to (see Figure 3), and that we therefore could not find a strong association between pitch and any motion feature. For example, the music excerpt *AladdinSane* featured a simple bass line with notes on the first and third beat, as well as a complex and dissonant piano solo. Participants reported in the short interviews that they stuck to the rhythm of the bass and tried not to become distracted by the piano. Regarding the vertical metaphor, contemporary dance movement does not resemble conducting an orchestra as performed by Truslit(Truslit, 1938). Ascending movement rather coincides with elevations of energy than elevations of pitch per se (cf. Van Dyck et al., 2013). These results are also in line with those of Wöllner and Hohagen (2017) and Burger et al. (2013) who both confirmed hand movement to be related to spectral flux of music or sound, rather than melodic direction. When participants cannot choose between rhythmic and melodic sound tracks, Kelkar and Jensenius (2018) showed that the "vertical metaphor", ascending melody corresponding to ascending movement in space, is just one out of six patterns to model melodic contour by movement. For less rhythmic music and musically trained persons (cf. Küssner et al. (2014)), it can be suitable to provide time series features that model less repetitive gestalts of music like melody. This was not accomplished by the motion features employed in this study.

For complexity that was predicted by irregular movement, sudden changes in the music were accompanied by abrupt movement (irregular rise and fall times) of participants. While moderate rhythmic complexity was shown to evoke a stronger urge to move according to Witek et al. (2014), high rhythmic complexity made participants uncomfortable to move to. The song rated most complex, *Le Sacre du Printemps* by

Stravinsky, made participants either stand still or give a performance ready for the stage depending on their dancing experience. For non-professional dancers (represented through the majority of participants), the moderate musical complexity seems to be especially stimulating.

Rhythmic entrainment works as a rather spontaneous and subconscious "synchronization" to the beat of the music. Therefore, rhythm-related properties of music might evoke more similar and repetitive movement among participants. However, the musical properties *complexity* and *pitch level + range* are stronger related to tonal and expressive qualities of music, and hence could describe more complex, conscious processes of music perception related to the participant's musical past experiences or whether or not a person identifies or "empathizes" with the artist (Egermann & McAdams, 2013). Thus, these properties of music might evoke less universal but more individual movement patterns among participants, and hence are more difficult to predict. Some participants reported that they did not want to move to certain excerpts because the music made them feel uncomfortable. Therefore, an expressive musical quality could also evoke motionlessness. We did not include those data samples into our evaluation[9], but it would be interesting to think about what we can learn from these observations for future work.

Besides, the missing link between melodic qualities of music and movement responses might also be explained by the findings of Gomez and Danuser (2007) who found that melody or pitch related properties of music did not appear to have any clear emotional or physiological association in general. Instead the musical properties *mode, rhythmic articulation, and harmonic complexity* were significant predictors for the experience of negative or positive valence. We cannot draw any conclusion for musical mode in our study because inter-rater reliability was not high enough for this property. The other two properties (*rhythmic articulation and harmonic complexity*) are inside of the rhythmicity and the complexity components. Thus, the poor performance of our model to predict pitch related musical properties are in line with those of Gomez and Danuser (2007). Another common observation between this and our study is that high arousal and physiological response were predicted by high rhythmic articulation (staccato), tempo (fast), and accentuation (marcato). These properties are represented in our *rhythmic* component, and partly as (tempo) in our *complexity* component. Furthermore, we assume that rhythmicity is stronger related to *arousal* whereas expressive properties are more related to *valence* which has been shown to be more difficult to predict by movement (cf. Amelynck, Grachten, van Noorden, and Leman, 2012, Camurri et al., 2004, or Irrgang and Egermann, 2016).

In this exploratory approach to embodied querying of music, we wanted to start with a broad range of music properties to test which properties are relevant for the danceability of music, and how they can be predicted by accelerometer-quantified movement. For future research, we are planning to focus on evaluating influences on music evoked movement of various rhythmic key properties like syncopation (cf. Sioros et al., 2014, Witek et al., 2014), microtiming or accentuation (cf. Naveda et al., 2011) in a modest but not simple complexity range.

[9] Music excerpts that made participants stand still as described in Section 4.2 were excluded from our evaluation

6. Conclusion

The presented study showed how smartphone-assessed accelerometer data can be used to quantify music-evoked movement. The findings provide a low-budget and mobile alternative to assess embodied music cognition, and hence offer to also carry out studies in the field for ecological validity. Furthermore, the suggested motion features are suitable to predict rhythm-related properties of music in a music recommendation scenario. Instead of sending text queries for music featuring a particular rhythm ("Samba", "Tango", "Hip Hop") or browsing playlists, one could use the most intuitive way to retrieve the desired music: perform the movement that the retrieved music should match. On the technical side, an extraction of motion features as time series might unravel additional links between movement and gestalt-like music features, and allow to further differentiate the experience on the one side and the querying of music on the other side. The observed movement preference of participants for rhythmic elements of music motivates further and more differentiated research in that domain. The relation between syncopation in music and movement deserves more research in particular. Concluding, the results confirm the feasibility of an embodied approach to Music Information Retrieval (even featuring very basic motion sensors) with focus on rhythmic properties of music.

Acknowledgement(s)

We want to thank Balthasar Effmert and Dmitry Grigoriev for their assistance in conducting the study.

Disclosure statement

The authors declare no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes on contributor(s)

Melanie Irrgang received a degree in computer science (M.Sc.) from TU Berlin and is a PhD student of the Audio Communication Group at TU Berlin. Her research interests are in the fields of movement and computing, machine learning and human-computer interaction. She is a bass and piano player, as much as a dancer.
Jochen Steffens is Professor of the Institute of Sound and Vibration Engineering at the Düsseldorf University of Applied Sciences with a Habilitation and a PhD in musicology and psychoacoustics and a formal background in audio and video engineering and psychology. He is interested in how we listen to sound and music, how we can predict and understand music listening behavior, and how listening affects our wellbeing as well as our emotions, perceptions, and behaviours. He is a piano and organ player as

well as a hobby jazz and swing dancer.

Hauke Egermann is Associate Professor of the Department of Music, University of York. He graduated in Systematic Musicology, Media Studies, and Communication Research (MA 2006, Hanover University for Music and Drama, Germany). Subsequently, he studied Neuroscience (PhD in Music Psychology/Neuroscience 2009, Center for Systems Neurosciences Hanover). He was Postdoctoral Research Fellow at the Centre for Interdisciplinary Research in Music Media and Technology (2009-2011, McGill University, Montreal, Canada). From 2011 to 2015, he lectured and researched at the Audio Communication Group (Technische Universität Berlin, Germany). In 2016, he was awarded his Habilitation in Musicology. Since 2016, he directs the York Music Psychology Group and conducts empirical research on production and experience of sound and music.

References

Amelynck, D., Grachten, M., van Noorden, L., & Leman, M. (2012). Toward E-Motion-Based Music Retrieval a Study of Affective Gesture Recognition. *IEEE Transactions on Affective Computing*, *3*(2), 250-259. Retrieved from `https://doi.org/10.1109/T-AFFC.2011.39`

Burger, B., London, J., Thompson, M. R., & Toiviainen, P. (2018). Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychological Research*, *82*(6), 1195–1211. Retrieved from `https://doi.org/10.1007/s00426-017-0894-2`

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2013). Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Frontiers in psychology*, *4*. Retrieved from `https://doi.org/10.3389/fpsyg.2013.00183`

Burger, B., Thompson, M. R., Luck, G., Saarikallio, S., & Toiviainen, P. (2014). Hunting for the beat in the body: on period and phase locking in music-induced movement. *Frontiers in human neuroscience*, *8*. Retrieved from `https://doi.org/10.3389/fnhum.2014.00903`

Camurri, A., Mazzarino, B., Ricchetti, M., Timmers, R., & Volpe, G. (2004). Multimodal analysis of expressive gesture in music and dance performances. In A. Camurri & G. Volpe (Eds.), *Gesture-based communication in human-computer interaction* (pp. 20–39). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from `https://doi.org/10.1007/978-3-540-24598-8`$_3$

Drake, C., Jones, M., & Baruch, C. (2000, 12). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, *77*, 251-88. Retrieved from `https://doi.org/10.1016/S0010-0277(00)00106-2`

Egermann, H., & McAdams, S. (2013). Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Perception: An Interdisciplinary Journal*, *31*(2), 139–156. Retrieved from `https://doi.org/10.1525/mp.2013.31.2.139`

Eitan, Z., Schupak, A., Gotler, A., & Marks, L., E. (2014). Lower Pitch Is Larger, Yet Falling Pitches Shrink - Interaction of Pitch Change and Size Change in Speeded Discrimination. *Experimental Psychology*, *61*, 273-284. Retrieved from `https://doi.org/10.1027/1618-3169/a000246`

Giordano, B., Egermann, H., & Bresin, R. (2014). The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity. *PlOS ONE*, *9*.

Godøy, R. I., Song, M., Nymoen, K., Haugen, M. R., & Jensenius, A. R. (2016). Exploring sound-motion similarity in musical experience. *Journal of New Music Research*, *45*(3), 210-222. Retrieved from `https://doi.org/10.1080/09298215.2016.1184689`

Gomez, P., & Danuser, B. (2007). Relationships Between Musical Structure and Psychophysiological Measures of Emotion. *Emotion*, *7*(2), 377–387. Retrieved from https://doi.org/10.1037/1528-3542.7.2.377

Grahn, J. A., & Rowe, J. B. (2009). Feeling the beat: Premotor and striatal interactions in musicians and nonmusicians during beat perception. *Journal of Neuroscience*, *29*(23), 7540–7548. Retrieved from https://doi.org/10.1523/JNEUROSCI.2018-08.2009

Greb, F., Steffens, J., & Schlotz, W. (2018). Understanding music-selection behavior via statistical learning: Using the percentile-lasso to identify the most important factors. *Music & Science*, *1*, 2059204318755950. Retrieved from https://doi.org/10.1177/2059204318755950

Honing, H. (2012). Without it no music: beat induction as a fundamental musical trait. *Annals of the New York Academy of Sciences*, *1252*(1), 85-91. Retrieved from https:doi.org/10.1111/j.1749-6632.2011.06402.x

Irrgang, M., & Egermann, H. (2016). From motion to emotion: Accelerometer data predict subjective experience of music. *PLOS ONE*, *11*(7), 1-20. Retrieved from https://doi.org/10.1371/journal.pone.0154360

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323–355.

Kelkar, T., & Jensenius, A. R. (2018). Analyzing free-hand sound tracings of melodic phrases. *Applied Sciences*, *8*. Retrieved from https://doi.org/10.3390/app8010135

Küssner, M. B., Tidhar, D., Prior, H. M., & Leech-Wilkinson, D. (2014). Musicians are more consistent: Gestural cross-modal mappings of pitch, loudness and tempo in real-time. *Frontiers in Psychology*, *5*. Retrieved from https://doi.org/10.3389/fpsyg.2014.00789

Leman, M. (2007). *Embodied Music Cognition and Mediation Technology.* London: MIT Press.

Leman, M., & Maes, P.-J. (2015). The role of embodiment in the perception of music. *Empirical Musicology Review*, *9*. Retrieved from https://doi.org/10.18061/emr.v9i3-4.4498

Luck, G., Saarikallio, S., Burger, B., Thompson, M., & Toiviainen, P. (2010). Effects of the big five and musical genre on music-induced movement. *Journal of Research in Personality*, *44*(6), 714 - 720. Retrieved from https://doi.org/10.1016/j.jrp.2010.10.001

Maes, P.-J., Leman, M., Palmer, C., & Wanderley, M. (2014). Action-based effects on music perception. *Frontiers in Psychology*, *4*. Retrieved from https://doi.org/10.3389/fpsyg.2013.01008

Moelants, D. (2002). Preferred tempo reconsidered. In C. Stevens, D. Burnham, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th international conference on music perception and cognition* (pp. 580–583).

Naveda, L. A., Gouyon, F., Guedes, C., & Leman, M. (2011). Microtiming Patterns and Interactions with Musical Properties in Samba Music. *Journal of New Music Research*, *40*(3), 225–238. Retrieved from https://doi.org/10.1080/09298215.2011.603833

Sievers, B., Polansky, L., Casey, M., & Wheatley, T. (2013). Music and movement share a dynamic structure that supports universal expression of emotion. *PNAS*, *110*(1), 70-75. Retrieved from https://doi.org/10.1073/pnas.1209023110

Sioros, G., Miron, M., Davies, M., Gouyon, F., & Madison, G. (2014). Syncopation creates the sensation of groove in synthesized music examples. *Frontiers in Psychology*, *5*. Retrieved from https://doi.org/10.3389/fpsyg.2014.01036

Styns, F., Van Noorden, L., Moelants, D., & Leman, M. (2007). Walking on Music. *Human Movement Science*, *26*, 297–305. Retrieved from http://dx.doi.org/10.1016/j.humov.2007.07.007

Toiviainen, P., Luck, G., & Thompson, M. R. (2010). Embodied Meter: Hierarchical Eigenmodes in Music-Induced Movement. *Music Perception: An Interdisciplinary Journal*, *28*(1), 59–70. Retrieved from https://doi.org/10.1525/mp.2010.28.1.59

Truslit, A. (1938). *Gestaltung und Bewegung in der Musik.* Friedrich Vieweg.

Van Dyck, E., Burger, B., & Orlandatou, K. (2017). The Communication of Emotions in Dance. In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge Companion to Embodied Music Interaction.* Routledge. Retrieved from https://doi.org/10.4324/9781315621364

Van Dyck, E., Moelants, D., Demey, M., Deweppe, A., & Leman, M. (2013). The impact of the bass drum on human dance movement. *Music Perception: An Interdisciplinary Journal*, *30*, 349-359. Retrieved from https:doi.org/10.1525/mp.2013.30.4.349

Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014, 04). Syncopation, body-movement and pleasure in groove music. *PLOS ONE*, *9*(4), 1-12. Retrieved from https://doi.org/10.1371/journal.pone.0094446

Wöllner, C., & Hohagen, J. (2017). Gestural qualities in music and outward bodily responses. In C. Wöllner (Ed.), *Body, sound and space in music and beyond: Multimodal explorations* (pp. 69–88). Routledge. Retrieved from https://doi.org/10.4324/9781315569628-5

Zatorre, R. J., Chen, J. L., & Penhune, V. (2007). When the brain plays music: audiotory-motor interactions in music perception and production. *Nat. Rev. Neurosci.*, *8*, 547-558. Retrieved from https://doi.org/10.1038/nrn2152